

Information Extraction (IE)

Computer-based analysis and visualization tools can assist in the interpretation of vast amounts of data, but the data must first be made accessible to those tools. The objective of Information Extraction (IE) technology is to provide computer software that automatically extracts and infers information from textual data sources and stores it into databases. Extracted information varies across three levels of complexity. Shallow Extraction refers to the extraction of simple types of information such as entities (the names of people, places, locations, etc.) numerical information (monetary values, percentages, etc.), and simple events.

Intermediate Extraction extracts inter-entity relationships (e.g., Person 2 works for Person 1) and more complex event information. Eventual research will progress to Deep Extraction, whereby information will be inferred, and complex relationships and event scenarios will be recognized. Developing domain-portable IE technology, capable of being customized for various applications across multiple domains, is an essential goal in order to shorten development times and reduce development costs.

